IED WORKING PAPER

A Note on the Use and Misuse of Summary Statistics in Regression Analysis

by

Robert V. Bishop

International Economics Division

Economics, Statistics, and Cooperatives Service
U.S. Department of Agriculture

A Note on the Use and Misuse of

Summary Statistics in Regression Analysis

by

Robert V. Bishop

September 22, 1980

Estimation Section; World Analysis Branch; International Economics Division; Economics, Statistics and Cooperatives Service; United States Department of Agriculture; Washington, D.C. 20250

Introduction

This paper discusses the effect of an autocorrelated error structure on the interpretation of traditional significance tests, in particular the t-test, F-test and R^2 measures. Emphasis is placed on first-order serial correlation 1/ which is a common and often serious problem encountered by researchers using time series data. 2/ This paper further suggests that even though many of the problems associated with an autocorrelated error structure are well known, many researchers ignore them and report results which range from being potentially misleading to grossly erroneous.

This paper surveys recent methodological developments concerning error structures which are "contaminated" with autocorrelation and draws implications that are relevant for the interpretation and application of empirical econometric research.

It is common to find instances where researchers report Durbin-Watson statistics 3/ that suggest an error structure that is first-order autocorrelated, but little if any account is taken for it. This paper discusses the bias introduced into the traditional tests of significance when autocorrelated errors are present, and suggests that interpretation of these tests under this condition is difficult if not impossible. Furthermore, it is asserted that the recognition of, and correction for first-order serial correlation via the usual Cochrane-Orcutt [1949] or Hildreth-Lu [1960] procedure is greatly inadequate in certain, although rather common situations.

If the error structure exhibits first-order autocorrelation, and we assume that no relevant variable has been ommited 4/, it is well known that the estimated regression coefficients are unbiased and consistent, but possess the

undesirable property of being inefficient. 5/ What is also well known in the literature, but often overlooked in practice, is that the usual tests of significance, when performed in the presence of autocorrelated errors, are biased. For example, if positive first-order autocorrelation is present in the error structure, the estimates of the standard errors on each of the coefficients (s_b) will be biased downward in most situations. When the standard error of the coefficient is underestimated, it is obvious that the t-statistic on that coefficient must be overstated since it is computed as $t = \frac{\hat{\beta}}{s_{\hat{\beta}}}$, implying greater explanatory power for that variable than really exists. This could easily lead to the inclusion of a statistically irrelevant variable into the final model. If the error structure exhibits negative serial correlation, the standard error of the coefficient is likely to be over-estimated, possibly leading to the elimination of a statistically significant variable from the model.

Granger and Newbold [1974] performed a series of tests in which the potential of discovering "spurious" relationships due to problems with autocorrelated errors was examined. They felt that a great deal of econometric work documented in the literature was permeated with "relationships" that existed only due to the failure of the researcher to remove autocorrelation from the error structure. They examined the values of the coefficient of determination [R^2] generated by regressions in a Monte Carlo experiment. Two independent series were generated, one being a random walk and the other a more complicated ARIMA (0,1,1) 6/ structure. They concluded that:

"It is quite clear from these simulations that if one's variables are random walks and one includes in regression equations variables which should in fact not be included, then it will be the rule rather than the exception to find spurious relationships. It is also clear that a high value for R^2 or $\bar{R}^2$ combined with a low value of d (Durbin-Watson Statistic) is no indication of a true relationship" 7/.

In a later paper, the authors elaborated their position which is quoted here at length:

> In time series regressions involving the levels of economic variables, one frequently sees coefficients of multiple correlation (R^2) much higher than 0.9. If these indicate anything at all, they presumably imply an extremely strong relationship between the dependent variable and the independent variables. This is extremely misleading on many occasions, as comments noting poor forecast performance which sometimes follow these equations will testify. In fact, the high R^2 values could be no more than a reflection of the fact that the dependent variable is highly autocorrelated and could easily be achieved simply by regression the variable on its own past. Thus, in such circumstances, the value of R^2 says nothing at all about the strength of the relationship between the dependent and independent variable. 8/

Summarizing thus far, if the error structure is first-order autoregressive [AR (1)] 9/, the OLS estimates of the regression parameters are: (a) Unbiased, (b) Consistent, but (c) Inefficient in small as well as large samples. The estimates of the standard errors of the coefficients in a model are biased downward if the residuals are positively autocorrelated and the independent variable itself is autocorrelated; and upward if they are negatively autocorrelated and the independent variable is again autocorrelated. Therefore, the calculated t-statistics is biased upward [downward] if the residuals are biased positively [negatively]. Granger and Newbold [1973] have demonstrated that the R^2 measure (both adjusted and unadjusted) 10/ is grossly misleading in the presence of an autocorrelated error structure. Granger and Newbold [1977] have further suggested that the regression results

can be defined as "nonsense" if the R^2 measure exceeds that computed for the Durbin-Watson statistic.

To demonstrate how misleading some regression statistics can be, an example is offered. Natural logarithms of quarterly measures of the U.S. price level (CPI) was regressed on the logarithm of the narrowly defined money stock (old M1) for U.S. over the period 1947-1978. This is a test of a model describing the Crude Quantity Theory of Money which states that changes in the exogenous money stock cause changes in the passive (endogenous) price level. The estimation results are presented in Table I and can be considered "nonsense results" as defined above. Table II presents regression results using the same model but employing a simple first-differencing 11/ transformation on the dependent and independent series. Letting M* denote the transformed money series, first differencing is accomplished as $M^*_t = M_t - M_{t-1}$. The Durbin-Watson statistic resulting from this estimation is still low, and a fourth-order 12/ autoregression was performed on the residuals. The results of this autoregression indicate that the error structure is more complex than first-order autoregressive due to significant coefficients on the lagged terms of lag greater than one. This example clearly demonstrates how sensitive the summary statistics of a regression can be to a first-

differencing transformation. This is important both in testing the theoretical specification of the model as well as in forming expectations of the forecasting ability of the model. For example, if one obtains and R^2 of .99 from a model estimated using levels, which may in reality only be explaining 30 percent of the variation of the dependent variable, one would not expect forecasts generated from this model to perform well. Therefore, choosing the appropriate forecasting model between one estimated using levels and another using changes based on R^2 only makes sense if the R^2 measure truly attests to the model's explanatory power. If the model is correctly specified, the choice wouldn't matter. Presentation of these results using raw data, and not presenting the Durbin-Watson statistic would be very deceptive at best and intellectually dishonest at worst. It must also be recognized that a "good" Durbin-Watson statistic is insufficient to conclude that the error structure is "contamination free" in terms of autocorrelation. This last point is alluded to when the results of the fourth-order autoregressions on the residuals are examined and is discussed below.

Methods of Correcting for Autocorrelation

Assuming that evidence of 1st order autocorrelation exists, one then would ask what can be done to correct for it, or if correction is appropriate. A rather simple (but often effective) approach suggested in the preceding section was to first-difference the data. This is equivalent to applying a filter (1- L) [where L is a lag operator such that $(1-L)X_t = X_t - X_{t-1}$] where p=1. This technique might be referred to alternatively as "pre-whitening" the input series, "filtering" the input series, or "applying a first-differencing transform" to the input series. Another possible method

differencing transformation. This is important both in testing the theoretical specification of the model as well as [illegible] the forecasting ability of the model. For example, [illegible] .99 from a model estimated using levels [illegible] explaining 99 percent of the variation [illegible] not expect forecasts generated from [illegible] choosing the appropriate [illegible] and another using [illegible] truly attests [illegible] specified the [illegible] raw data, and [illegible] deceptive [illegible] recognized that [illegible] that the [illegible] This test [illegible] autoregressi[illegible]

Methods

[illegible]

Table I--Regression Results of $\ln(P_t) = f[\ln(M1_t)]$

Variable	Est Coefficient	t-statistic
Intercept	-0.2878	1.42
ln M1	0.9686	21.92 *
Linear Trend	-0.0014	3.64 *

R^2 =.98 F = 6027
DW =.06 SEE = .177137

Table II--Regression Results of $[\ln(P_t)-\ln(P_{t-1})] = f[\ln(M1_t)-\ln(M1_{t-1})]$ a/

Variable	Est. Coefficient	t-statistic
Intercept	0.0024	1.61
$\ln M1_t - \ln M1_{t-1}$	0.1824	1.53
Linear Trend	0.0001	3.47*

R^2 = .15 F= 21.7
DW = .69 SEE = .0080963

Where: R^2 = corrected R^2
DW = Durbin Watson Statistic
SEE = Standard error of the estimate
F = Computed F-statistic

a/ Approximates: % change in price = f(% change in M1)
* significant at the 05. level

for eliminating autocorrelation from the error structure lies in "quasi-differencing" or applying the filter $(1-\rho L)$ where $-1 < \rho < 1$. For example, if ρ is assumed to equal .75, the application of the filter (1-.75L) to Y_t results in a transformed series Y_t^* which is equal to $Y_t-(.75)Y_{t-1}$. The Cochrane-Orcutt Iterative Technique 13/ estimates a value of ρ rather than assumming it using the residuals computed from a regression of the untransformed or "raw" data. A potential problem with this technique is that the value of ρ can be chosen that satisfies the selection criterion of the computer algorithm but does not purge the error structure of first-order autocorrelation. This occurs when the estimation converges to a value that is a local rather than global minimization of the sum of squared residuals. Another problem arises if forecasts are generated from a model

estimated with this technique since (1) the coefficient on the intercept on the must be corrected, and (2) errors in the forecast will tend to compound over time. 14/

Another technique available to the researcher is the Hildreth-Lu procedure which uses values of ρ prechosen by the researcher. This technique is more robust against converging to an inappropriate value of ρ but is still vulnerable to the two criticisms leveled above if estimates derived from it are used for forecasting.

Further Discussion of the Analysis of Regression Residuals

Referring to the treatment of residuals in econometric models, Granger and Newbold (1977) state that "The traditional approach has been to assume first the residuals to be white noise 15/ and to check this assumption by looking at the Durbin-Watson statistic, which effectively measures first-order serial correlation of the residuals. If evidence of significant first-order serial correlation is found, the residuals are assumed to be first-order autogressive ... there is little reason to suppose that the correct model for residuals is AR(1); in fact, if the variables involved are aggregates and involve measurement error, an ARMA 16/ model is much more likely to be correct". 17/ If the errors are characterized by a mixed moving average autoregressive structure, time series modeling of the residuals can be employed. Application of this technique is beyond the scope of this paper, but the researcher should be aware of this powerful and innovative approach. 18/

Wallis (1972) found that models which use quarterly data can often be plagued by fourth-order autocorrelation either when seasonally adjusted or

unadjusted data have been used. He suggests that monthly or weekly models may also have a seasonal component remaining that can manifest itself as either 12th or 52nd order autocorrelation in the error structure respectively. We would not suggest that twelfth or fifty-second order prefilters be constructed and applied to the data, but the residuals can be modeled using the methodology of Box and Jenkins [1976] as mentioned above, employing seasonal differencing to the residuals and then estimating the order of the autoregressive and moving average components of the characterization. Again, full discussion of this technique is beyond this paper's scope.

Pierce [1977] examines the issue of complex error structures in detail emphasizing the skepticism that experienced researchers exhibit when confronted with high R^2 measures. He notes that the R^2 measure is properly constructed as a measure of effects between variables, while in many applications, the measure is contaminated by within variable effects. This is seen when lagged values in a relationship are combined with serial correlation in the dependent variable which "means that part of variance of y (the dependent variable) is explainable by its own past. This R^2 ... will generally include effects attributable simply to lagged values of y." 19/ For this reason, Pierce asserts that the estimated R^2's using time series data exhibit a great deal of sensitivity to "prefiltering" which removes this within variable effect. This was discussed above and the example presented in this paper should serve to underscore this important point. 20/ Discussing the great difference between the apparent contribution to R^2 made by a lagged dependent variable expressed as a level (untransformed lagged dependent variable) and the contribution of that variable expressed as changes (i.e. 1st

unadjusted data have been used. He suggests that monthly [illegible]
also have a seasonal component remaining that is reflected [illegible]
12th or 52nd order autocorrelations in the [illegible]
would not suggest [illegible]
and applied [illegible]
methodology [illegible]
differencing [illegible]
autoregressive [illegible]
(vii) discussion [illegible]

Pierce [illegible]
emphasizing [illegible]
information [illegible]
properly [illegible]
application [illegible]
test [illegible]
in the [illegible]
dependent [illegible]
[illegible]

differenced data) prompted Pierce to state "This phenomenon results in an intrinsic ambiquity in conventional R^2 measures, and it is perhaps this ambiguity which underlies the rather limited faith often accorded these measures by persons experienced with time series data." 21/

Conclusion

The researcher is faced with a great deal of difficulty in the interpretation of the results from his modeling effort. Measures computed from summary statistics such as the coefficient of variation of the dependent variable (standard error of the regression divided by the mean of the dependent series) are potentially meaningless. The explanatory power of the model may be grossly over or under stated. The variables included in the model might be irrelevent, or statistically significant variables may be excluded due to biased t-statistics. To protect oneself from being mislead by the results of one`s research, the error structure must be critically examined. The Durbin-Watson statistic should always be examined 22/ as a bare minimum, and at the same time, the researcher must be aware that this is not necessarily sufficient to insure a meaningful interpretation of his results. Additional work in residual analysis is continuing with the employment of sophisticated techniques such as time series modeling (using the methodology of Box and Jenkins [1970]) as well as analysis in the frequency domain via spectral techniques. At this point there are no easy answers, but there are a great deal of complex issues which researchers must address themselves to if the fruit of their research is to be useful for policy makers and other researchers in that area.

ENDNOTES

1/ First order serial correlation implies that the error term in period t is correlated with the error term in period (t-1) i.e.:

$$e_t = f(e_{t-1})$$

2/ Autocorrelation is, by definition, serial correlation in time series data. Serial correlation can also exist in cross-sectional data such as spatial correlation (across geographic regions) or mutual correlation (across groups such as income).

3/ The Durbin-Watson statistic (d) is computed as: $d = \frac{\Sigma(\hat{u}_i - \hat{u}_{i-1})^2}{\Sigma u_i^2}$

This approximates $d=2(1-\rho)$ where ρ is the estimated 1st order autocorrelation coefficient of the residuals in the model. As $\rho \to 0$, d approaches two. By examining a table for upper (d_u) and lower (d_l) bounds on this statistic for the appropriate degrees of freedom the null hypothesis of no autocorrelation can be tested.

4/ If a relevant causal variable has been omitted, the estimates are also biased.

5/ The properties of estimators in the presence autocorrelation are contained in many good Econometrics Texts, one such example is Kmenta's Elements of Econometrics (1971) pps. 273-282.

6/ A process characterized as ARIMA (0,1,1) is an integrated (regular differencing for stationarity is applied) mixed autoregressive moving average model. ARIMA (0,1,1) implies no autoregressive component, one moving average component, and one level of regular differencing. A good, though incomplete, introduction to time series modeling is contained in Hammond, E., "Forecasting the Many Stock with Time Series Models".

7/ "Spurious Regressions in Econometrics" Journal of Econometrics 2, pps. 111-120.

8/ "The Time Series Approach to Model Building". In New Methods in Business Cycle Research: Proceedings From a Conference Federal Reserve Board of Minneappolis. 1977 pps. 12-13.

9/ An error structure that is AR (1) is one that exhibits only simple first-order autocorrelation.

10/ Adjusted R^2 or $\bar{R}^2 = 1 - (1-R^2)\{(T-1)/(T-K)\}$

where T= number of observations and K= number of estimated parameters. This adjustment is for degrees of freedom in the estimating equation lost with the inclusion of additional variables. This adjustment offsets the upward bias in the unadjusted R^2 which is most dramatic with a small sample size.

11/ First-differencing of natural logarithms approximates a percentage rate of change.

12/ $e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \rho_3 e_{t-3} + \rho_4 e_{t-4}$

13/ Initial and subsequent values of p are estimated via an autoregression of the residuals (first-order).

i.e: $e_t = pe_{t-1} + u_t$

14 This is seen when we consider a forecast one period into the future for some variable Y:

$$Y_{t+1} = a_1(1-\rho) + a_2(X_{t+1} - \rho X_t) + \rho Y_t$$

if Y_{t+1} is understated, Y_{t+2} will also be understated since Y_{t+1} will enter the computation on the right hand side.

15/ White noise implies that all of the non-random components have been removed from the series, and no additional "information" remains.

16/ Autoregressive Moving Average model needing no regular differencing.

17/ Granger and Newbold (1977), p. 9.

18/ This procedure attempts to explain the effect of variables which had been excluded from the model. One can argue that the parameter estimates thus obtained are more "proper". The forecasting performance is encouraging.

19/ "R^2 Measures for Time Series," Special Studies Paper #93, Federal Reserve Board of Governors, 1978, p.3

20/ The example in this paper did not include a lagged dependent variable, however, the sensitivity of our examples to prefiltering is obvious.

21/ Pierce (1978) ibid.

22/ If a lagged dependent variable is included, the Durbin-Watson statistic is biased toward two and is therefore more difficult to interpret. If it is very low (near 0) or very high (near 4), the residuals are most likely autocorrelated. If the sample size is large, one can construct a modified h-statistic to test for autocorrelation in the presence of a lagged dependent variable used as an explanatory variable. Reference for this is found in Durbin, J, "Testing for Serial Correlation in Least Squares Regression when some of the Regressors are Lagged Dependent Variables" Econometrica (May 1970) Vol. 38 No. 3, pp. 410-421.

Bibliography

Berndt, E.R. and Savin, N.E., 1979, "Estimation and Hypothesis Testing in Singular Equation Systems with Autoregressive Disturbances," Econometrica Vol. 43, No. 5-6 (November), pps. 937-957.

Box, G.E.P. and Jenkins, G.M., 1976, Time Series Analysis (San Fran.: Holden-Day).

Box, G.E.P. and Pierce, D.A., 1970, "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models", Journal of the American Statistical Association 65, December, pps. 1509-1526.

Cochrane, D. and Orcutt, G. H. "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms" Journal of the American Statistical Association, vol. 44, No. 245 (March 1949); pps. 32-61.

Durbin, J., 1970, "Testing for Serial Correlation in Least Squares Regression When Some of the Regressors are Lagged Dependent Variables," Econometrica 38, May, pps. 410-421.

Engle, Robert F., 1974, "Specification of the Disturbance for Efficient Estimation," Econometrica 42, January, pps. 135-146.

Fuller, W.A. and Martin J.E. "The Effects of Autocorrelated Errors on the Statistical Estimation of Distributed Lag Models," Journal of Farm Economics, pps. 71-82.

Hammond, Elizabeth, 1977, "Forecasting the Money Stock with Time Series Models", Federal Reserve Board of New York Research Paper #7729 (May).

Kmenta, Jan, 1971, Elements of Econometrics, (New York: Macmillan).

Granger, C.W.J. and Morris, M.J., 1976, "Time Series Modelling and Interpretation," Journal of the Royal Statistical Society A, 139, Pt. 2, pps. 246-257.

Granger, C.W.J. and Newbold, P., 1974, "Spurious Regressions in Econometrics", Journal of Econometrics 2, pps. 111-120.

Granger, C.W.J. and Newbold, P., 1977, "The Time Series Approach to Econometric Model Building", New Methods in Business Cycle Research: Proceedings from a Conference, Federal Reserve Bank of Minneapolis, pps. 7-22.

Griliches, Zvi, 1967, "Distributed Lags: A Survey," Econometrica 35, pps. 16-49.

Henshaw, R.C. Jr., 1966, "Testing Single-Equation Least Squares Regression Models for Autocorrelated Disturbances," Econometrica 34, July, pps. 646-660.

Hildreth, C. and Lu, J.Y., 1960, "Demand Relations with Autocorrelated Disturbances" Technical Bulletin 276, Michigan State Univerisity, Agricultural Experiment Station.

McLeod, A.I., 1979, "Distribution of the Residual Cross-Correlation in Univariate ARMA Time Series Models," Journal of the American Statistical Association 74, December, pps. 849-855.

Maddala, G.S., 1977, Econometrics, (New York: McGraw Hill).

Maddala, G.S. and Rao, A.S., 1973, "Tests for Serial Correlation in Regression Models With Lagged Dependent Variables and Serially Correlated Errors," Econometrica 41, July, pps. 761-774.

Nelson, C.R., 1973, Applied Time Series Analysis for Managerial Forecasting (San Fran.: Holden-Day).

Nerlove, M. and Wallis, K., 1966, "Use of the Durbin-Watson Statistic in Inappropriate Situations," Econometrica 34, January, pps. 235-238.

Pierce, David A., 1978, "R^2 Measures For Time Series," Special Studies Papers #93, Federal Reserve Board of Govenors, October.

Rao, P. and Griliches, Zvi, 1969, "Small-Sample Properties of Several Two-Stage Regression Methods in the Context of Autocorrelated Errors," Journal of the American Statistical Association, March, pps. 253-272.

Saracoglu, R., 1977, "The Maximum Likelihood Extimation of Parameters in Mixed Autoregressive Moving-Average Multivariate Models," Staff Paper #20, Federal Reserve Bank of Minneapolis, (April).

Theil, H. and Clements, K., 1979, "Recent Methodological Advances in Economic Equation Systems," Unpublished.

Tillman, J.A., 1975, "The Power of the Durbin Test," Econometrica 43, November, pps. 959-974.